# Characterisation of pQTLs for 91 proteins

Authors to be added here

## 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](mailto:jp549@medschl.cam.ac.uk)).

***Last updated 26/3/2019***

## Abstract

BACKGROUND. METHODS. FINDINGS. INTERPRETATION.

## Introduction

Proteins involve in many biological processes and large-scale protein-wide genomic analysis (PGWAS) can provide important information on quantitative trait loci associated with proteins (pQTLs) as highlighted in recent work by Sun et al. (2018) while many aspects linking proteins remain unclear but could only be revealed via large-scale studies.

We conducted analysis with Olink/INF1 panel, consisting of 91 proteins, using cohorts within the SCAndinavian coLLaboration for Olink plasma Protein genetics (SCALLOP) consortium, for discovery and follow-up of genetic associations with proteins on the Olink Proteomics platform.

We identified pQTLs, the mechanistic basis of their cis- and trans- effects by functional annotation, pleiotropic effects, as well as causal role on disease outcomes such as CHD and other downstream analysis.

## Data and analysis

### The SCALLOP/INF1 consortium

The contributing cohorts in this study are listed in Supplementary Table S1. The original Olink INFlammation panel contained 92 proteins whose raw measurements, including those beyond lower limit of detection, were subject to a rank-based inverse normal transformation. Quality control on cohort level involves SNP/sample call rates, gender mismatch, abnormal inbreeding coefficient, failed cryptic relatedness test, ancestry outlier, heterozygosity and Hardy-Weinberg equilibrium test. 1000 genomes imputation, build 37 (hg19) positions. At the time of analysis, BDNF was dropped from the panel so will not be analysed.

### Association analysis

Multiple linear regression was conducted on protein data and genotypes as well as sex, age, principal components and other cohort specific covariates under an additive genetic model. As individual level data from KORA study rather than GWAS summary statistics were available, the protein normalization and association testing were done centrally.

Software which account for genotype uncertainty, such as SNPTEST were used but due to the relatively large number of proteins, results based on PLINK were also accepted, both amended with outputs from qctool –snp-stats.

## Meta-analysis

R package QCGWAS was used to generate cohort-level Q-Q and Manhattan plots, suggesting appropriate MAF cut-off. Meta-analysis were performed centrally using the inverse-variance weighted analysis of regression betas and standard errors, as implemented in the software METAL, version 28.8.2018. Filters such as MAF, HWE, minimum sample size and imputation quality filters were applied at the meta-analysis. Genomic control was not considered on the cohort level. Q-Q and Manhattan plots were generated from R package qqman while regional association plot from LocusZoom 1.4. The Bonferroni threshold for the genome-wide analyses is approximately 5 x 10-10. Independent loci were assessed by PLINK and GCTA and cis/trans classifications were obtained using customised R functions. The reference panels for these analyses included 1000Genomes release 3 as well as INTERVAL-based UK10K+1000Genomes. Variants explained were approximated with where *T* is the total number of variants, the chi-squared statistic and the associate sample size, respectively (Giri, et al. 2019).

PhenoScanner was used for variant annotation. Positive controls were applied on OPG and TNFSF14. In particular, findings on OPG are in line with earlier report (Kwan et al. 2014).

The GWAS summary statistics were further used in finemapping experiment via several software including finemap and JAM using approximately independent LD blocks, as well as gene enrichment and pathway analysis.

## Finemapping

Attempt was made through approaches implemented in several software, including PLINK, GCTA and finemap.

## Pathway analysis

Attempt was also made for based on cis signals.

## Heritability analysis

Individual level data from the INTERVAL study were analysed with GCTA, to be followed by counterpart for GWAS summary statistics.

## Transcriptomewide association analysis

This mirrors work by Mancuso et al. (2017) and work on mQTL by McRae et al. (2018) and Yengo, et al. (2018).

## Results

The list of variants is shown in Tables xx. The Q-Q, Manhattan, LocusZoom, and forest plots are shown in Supplementary Figures xx. The cis/trans classification is shown in Supplementary Table x along with chord diagram in Figure.

## Discussions

We were able to assess the protein-genetic associations in the OLINK panel. The variants identified were further characterized on cis/trans effects, evidence of pleiotropy, utility as instruments for causal inference through Mendelian randomization, as well as biological pathways. By analogy to polygenic score and protein score (Ganz et al. 2016), an inflammation score could also be built.

We were also be able to develop relevant functions in Bash as well as R, some of which were made generic through R package gap.

## Acknowledgements

We wish to thank participants from the SCALLOP studies and collaboration from colleagues for making this work possible.

## References

Folkersen L, et al. (2017). Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. PLoS Genetics 13(4), doi.org/10.1371/journal.pgen.1006706.

Ganz P, Heidecker B, Hveem K, Jonasson C, Kato S, Segal MR, Sterling DG, Williams SA (2016). Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. *JAMA* 315(23):2532-41. doi: 10.1001/jama.2016.5951.

Giri A, et al. (2019). Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat Genet* 51:51-62.

Kwan JSH, et al. (2014). Meta-analysis of genome-wide association studies identiﬁes two loci associated with circulating osteoprotegerin levels. Hum Mol Genet 23(24): 6684–669.

Mancuso N, et al. (2017). Integrating gene expression with summary association statistics to identify susceptibility genes for 30 complex traits. Am J Hum Genet 100:473-487

McRae AF, et al. (2018). Identification of 55,000 Replicated DNA Methylation QTL. *Sci Rep.* 8(1):17605. doi: 10.1038/s41598-018-35871-w.

Sun BB, et al. (2018). Genomic atlas of the human plasma proteome. Nature 558: 73–79.

Yengo L, et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Hum Mol Genet*. 27(20):3641-3649. doi: 10.1093/hmg/ddy271.

## Supplementary information

## URLs

Additional information about this investigation is available from GitHub, <https://github.com/jinghuazhao/INF>.

URL: SCALLOP, <https://www.olink.com/scallop/>; METAL <https://github.com/statgen/METAL>; PLINK, <http://zzz.bwh.harvard.edu/plink>; KING, <http://people.virginia.edu/~wc9c/KING/>; LDetect-data (for approximately independent LD blocks), <https://bitbucket.org/nygcresearch/ldetect-data>; FUSION, <http://gusevlab.org/projects/fusion/>, LocusZoom, <https://github.com/statgen/locuszoom-standalone>; R. <https://cran.r-project.org>; Uniprot, [https://www.uniprot.org](https://www.uniprot.org/):

## Cohort information

**BioFinder**.

**EGCUT**.

**INTERVAL.** The INTERVAL study: NHSBT (11-01-GEN) and the NIHR-BTRU in Donor Health and Genomics (NIHR BTRU-2014-10024) at the University of Cambridge in partnership with NHSBT. This study was partially funded by Merck. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health of England, or NHSBT. The Cardiovascular Epidemiology Unit at the University of Cambridge: UK MRC (G0800270), BHF (SP/09/002), UK NIHR Cambridge Biomedical Research Centre, ERC (268834), and European Commission Framework Programme 7 (HEALTH-F2-2012-279233).

See Astle, et al. (2016), Sun, et al. (2018).

**KORA**. The KORA study was initiated and financed by the Helmholtz Zentrum München – German

Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ and by the Competence Network Asthma and COPD (ASCONET), network COSYCONET (subproject 2, BMBF FKZ 01GI0882) funded by the German Federal Ministry of Education and Research (BMBF).

Genotypes were available on Affymetrix Axiom chips on 3788 individuals. The imputation was done through IMPUTE 2.3.2 with 1000Genomes phase 3 reference panel. A total of 1,070 individuals with both genotypes and protein data were used in association analysis via SNPTEST 2.5.2, where the per-sample missing proportion was obtained from qctool 2.0.1, and covariates were sex, age, and five PCs from GCTA 1.91.7beta. IBD information was obtained via KING 2.1.6. The total number of variants were 81,651,446.

**MadCam**. The MadCam trial samples were all from baseline but the patients included were moderate ulcerative cholitis. Several details on the trial are available from [Vermeire S](https://www.ncbi.nlm.nih.gov/pubmed/?term=Vermeire%20S%5BAuthor%5D&cauthor=true&cauthor_uid=28527704), et al. (2017).

**NSPHS**. Swedish Medical Research Council (K2007-66X-20270-01-3, 2011-5252, 2012-2884 and 2011-2354), the Foundation for Strategic Research (SSF). NSPHS as part of European Special Populations Research Network (EUROSPAN) was also supported by the European Commission FP6 STRP (01947, LSHG-CT-2006-01947).

**ORCADES**. Supported by the Chief Scientist Office of the Scottish Government (CZB/4/276, CZB/4/710), the Royal Society, the MRC Human Genetics Unit, Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947). ORCADES DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh.

Imputation panel was HRC.

**RECOMBINE**.

**STABILITY**.

**STANLEY**. The study consists of lah1 and swe6 subcohorts.

**VIS**. CROATIA-Vis is a family-based, cross-sectional study in the isolated island of Vis, Croatia that included 1,056 examinees aged 18-93. It is a genetic epidemiology study that aims to discover genetic factors that influence traits (e.g. height) or the risk of common complex diseases. The cohort is very well characterised with detailed phenotyping and genotyping information available.

Imputation panel was HRC.

## References

[Vermeire S](https://www.ncbi.nlm.nih.gov/pubmed/?term=Vermeire%20S%5BAuthor%5D&cauthor=true&cauthor_uid=28527704), et al. (2017). Anti-MAdCAM antibody (PF-00547659) for ulcerative colitis (TURANDOT): a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet* 390(10090):135-144. doi: 10.1016/S0140-6736(17)30930-3.

## Supplementary tables

S1. Study information

S2. Signals from the INTERVAL study

S3. cis/trans classification from the INTERVAL study

## Q-Q and Manhattan plots

## Regional association plots

## Forest plots

## Pathway analysis

## PheWAS results

## EWAS results