# Genomic analysis of 91 Olink/inflammation 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).

***Last updated 12/4/2019***

## Abstract

BACKGROUND. METHODS. FINDINGS. INTERPRETATION.

## Introduction

Proteins involve in many biological processes and much can be learnt about these and their relationship with diseases from and large-scale protein-wide genomic studies (PGWAS) with respect to the so-called quantitative trait loci with proteins (pQTLs), as highlighted in recent work by Sun et al. (2018), for which it is difficult to know otherwise.

In contrast to the a broad landscape of human plasma proteome as reported in Sun et al. (2018) based on the SomaLogic platform in the INTERVAL study, our focus here is on proteins measured on specific functions in the inflammation panel of Olink originally consisting of 92 proteins. Moreover, to gain more statistical power we also assembled data from a number of other cohorts within the SCAndinavian coLLaboration for Olink plasma Protein genetics (SCALLOP) consortium, including 966 individuals with sequence data.

In the following we report our findings, particularly on pQTLs, their their cis- and trans- effects through functional annotation, pleiotropic effects, as well as causal role on disease outcomes such as CHD and other downstream analysis.

## 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 identify and validate protein-specific genetic associations in the OLINK/INF panel, which were further characterised with respect to their cis/trans effects, pleiotropic roles, 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.

## Methods

## Olink/INF panel

The inflammation panel is among many focused on a specific area of disease, targeting 92 established and/or exploratory biomarkers.

### The INTERVAL study

The INTERVAL study was designed (Moore et al. 2014).

### The SCALLOP/INF1 consortium

The contributing cohorts in this study are listed in Supplementary Table studies. The 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

Prior to the meta-analysis, extensive effort were paid on quality control of the GWAS summary statistics with available information such as MAF, HWE, and imputation score. To facilitate this, cohort-level Q-Q and Manhattan plots were generated with R package qqman and QCGWAS. Meta-analysis were performed using the inverse-variance weighted analysis of regression betas and standard errors, as implemented in the software METAL, version 28.8.2018 and the results were additionally visualized with regional association plots from LocusZoom 1.4. Independent loci were assessed by PLINK clumping, using Bonferroni threshold for the genome-wide analyses approximately 5 x 10-10, 1000Genomes release 3 data as reference panel, and different degrees of linkage disequilibrium. cis/trans classifications were obtained using customised bash and R functions. A similar analysis was performed via GCTA. The above were done iteratively to ensure validity of the findings. Eventually, the reference panels for these analyses was INTERVAL-based imputed genotypes from UK10K+1000Genomes reference panel, and approximately independent LD-blocks were used. Genomic heritability was also assessed with GCTA. 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).

## Acknowledgements and funding information

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

**CROATIA-Korcula/Split/Vis:** MRC Human Genetics Unit core funding, the Ministry of Science, Education and Sport in the Republic of Croatia (216-1080315-0302) and the Croatian Science Foundation (grant 8875) (funding to I. Rudan, C. Hayward, S.M. Kerr, O. Polasek, V. Vitart, and J. Marten).

**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).

**KORA F4 and KORA S3:** 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).

**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.

## 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 is a prospective cohort study of approximately 50,000 participants of mostly European ancestry, nested within a pragmatic randomized trial of blood donors. Between 2012 and 2014, blood donors 18 years and older were consented and recruited from 25 NHSBT (National Health Service Blood and Transplant) static donor centers across England. Participants are predominantly healthy individuals since people with major disease (myocardial infarction, stroke, cancer etc.) are ineligible for donation, as are those who report being unwell or having had recent illness or infection. Participants completed online questionnaires containing basic lifestyle and health-related information, including self-reported height and weight, ethnicity, current smoking status, alcohol consumption, doctor-diagnosed anemia, use of medications (hormone replacement therapy, iron supplements) and menopausal status. The INTERVAL study was approved by the Cambridge (East) Research Ethics Committee and UK Biobank was approved by the North West Multi-center Research Ethics Committee (MREC). Informed consent was obtained from all participants.

See Moore, et al. (2014), Astle, et al. (2016), Sun, et al. (2018).

The KORA studies (Cooperative Health Research in the Region of Augsburg) are a series of independent population based studies from the general population living in the region of Augsburg, Southern Germany. The **KORA S3** study including 4,856 individuals was conducted in 1994/95. Spirometry was measured during a follow up in 1997/98 for all participants younger than 60 years who did not smoke or use inhalers one hour before the test. **KORA F4** including 3,080 individuals was conducted from 2006-2008 as a follow-up study to KORA S4 (1999-2001). 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).

**Northern Sweden Population Health Study** (NSPHS) represents a cross-sectional study conducted in the communities of Karesuando (samples gathered in 2006) and Soppero (2009) in the subarctic region of the County of Norrbotten, Sweden. Spirometry was performed in sitting position without noseclips using a MicroMedicalSpida 5 spirometer (http://www. medisave.co.uk). Three consecutive 28 lung function measurements per participant were done and the maximum value per measured lung function parameter was used for further analysis. Relatedness was taken into account by applying the "polygenic" linear mixed effects model. Genome-wide association analysis was performed using a score test, a family-based association test27 which uses the residuals and the variance-covariance matrix from the polygenic model and the SNP fixed effect coded under an additive model.

**Orkney Complex Disease Study** (ORCADES) is an ongoing family-based, cross-sectional study in the isolated Scottish archipelago of Orkney. Spirometry was performed in the sitting position without nose clips, using a Spida handheld spirometer. Measurements were repeated once and the better reading was used for analysis.

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

Studies. Study information

INT1. Signals from the INTERVAL study for r2=0 clumping

INT2. cis/trans classification on INT1

INT3. Signals from the INTERVAL study for r2=0.1 clumping

INT4. cis/trans classification on INT3

INF1.snp-stats. Basic summary statistics by cohort

INF1. Signals from the SCALLOP/INF1 for r2=0 clumping

INF2. cis/trans classification on INF1

INF3. Signals from SCALLOP/INF1 for r2=0.1 clumping

INF4. cis/trans classification on INF3

INF5. Signals from SCALLOP/INF1 for r2=0.1 cojo analysis

INF6. cis/trans classification on INF5

## Q-Q and Manhattan plots

## Regional association plots

## Forest plots

## Pathway analysis

## PheWAS results

## EWAS results