# CVD1/INF1 protein-HF analysis

Included here is data on the 26 overlapping proteins (no information on IL-4) from the Olink cvd1 & inf1 panels.

Steps to set up the environment are outlined in [notes](notes/README.md), while MendelianRandomization v0.6.0 is used together with a bug fix in workflow/scripts/MR\_functions.R. The directory input/ can potentially be built from rules defined in the workflow.

module load miniconda3/4.5.1  
export mypath=${HOME}/COVID-19/miniconda37  
conda activate ${mypath}  
# A dry run of the workflow.  
snakemake -n  
# Analysis (no --use-conda option since all R packages are more up-to-date locally)  
snakemake -c all

which gives output/MR.csv (MR results) and Obs.csv (observational results).

Some related operations are also ready.

snakemake --dag | \  
dot -Tpdf > dag.pdf  
snakemake --rulegraph | \  
dot -Tpdf > rulegraph.pdf  
report --report report.html

This MarkDown document is obtained via Rscript -e 'knitr::knit("README.Rmd")'.

|  |  |  |  |
| --- | --- | --- | --- |
| MR assumption | Key strength of cis- pQTL instruments for protein exposures | Potential bias | Study design consideration to minimise bias |

Relevance:  
genetic variants are associated with the exposure of interest | cis-pQTL variants derived from genome-wide association studies using circulating protein data are associated with expression of protein of interest by definition | \* Spurious pQTL associations relating to assay binding (cross reactivity, epitope effects) \* Limited abundance of protein in plasma or limited assay affinity leading to suboptimal pQTL detection \* Limited contribution of cis-variant to plasma protein variance in population | \* Use of largest available pQTL data for the set of proteins of interest from GWAS meta-analysis of circulating proteins \* Protein-level quality control based on limit of detection \* Instrument selection based on strength of association with circulating protein levels \* Use of multi-instrument MR model to improve statistical power | Independence:  
genetic variants is not associated with confounders between exposure and outcome of interests | It is unlikely that conventional confounding factors between protein and outcome (e.g. expression of other proteins, risk factors affecting both levels of protein and outcome of interest) can affect genetic variation | \* Confounding by population structure \* High linkage disequilibrium between selected cis-pQTL instrument variants and coding or regulatory variants affecting the expression or function of other proteins \* Overlapping gene regions | \* Adjustment for population structure \* Instrument pruning based on linkage disequilibrium metrics \* Instrument is selected from ±200 kb flanking region of protein-encoding genes to limit contamination by variants with structural gene | Exclusion restriction:  
Genetic variants affect the outcome only through effects on exposure (no horizontal pleiotropy) | The central dogma of molecular biology denotes that the functional form of protein is expressed through transcription of cognate gene and translation machinery, implicating that any effect of cis-pQTL variants on outcome is downstream of its effect on the protein of interest | \* Alternatively spliced gene resulting in the expression of additional distinct proteins, other than the protein of interest, with effects on the outcome \* Selected cis-pQTL instruments affects the expression or function of a microRNA that regulates the translation of transcripts from multiple other genes | \* Multiverse sensitivity analysis with permutation of instrument selection parameters and MR models to test robustness of the main MR estimates that might arise due to presence of invalid cis-pQTL instruments