Kaur Alasoo: Implement step-wise conditional eQTL meta-analysis with regenie.

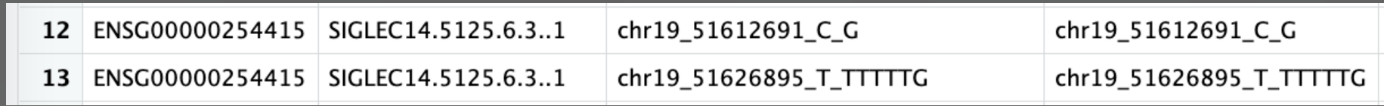
We know that many genetic loci associated with human traits contain multiple independent causal variants. The effects of these independent loci can be disentangled using statistical fine mapping approaches such as the Sum of Single Effects model. However, statistical fine mapping can be reliably performed only on homogeneous datasets using a single genotyping technology. Using statistical fine mapping in a meta-analysis setting has been demonstrated to be unreliable.

A less accurate alternative to meta-analysis statistical fine mapping than should produce less false positives is conditional meta-analysis. In this setting, multiple studies are first meta-analysed to identify the most strongly associated variant at each locus. This variant is then regressed out from all individual datasets and the resulting summary statistics are again meta-analysed. This process is repeated until some predefined criteria is met (e.g. no more significant results).

In this project, you will use some open-access eQTL datasets to implement conditional meta-analysis using the regenie or tensorqtl software (e.g. pseudo-bulk data from OneK1K + Randolph\_2021 or GEUVADIS LCLs). You will test the workflow by performing colocalization analysis with the protein QTLs from the INTERVAL cohort.

If the workflow works well, then it can later be extended to multiple datasets in the eQTL Catalogue (skin, muscle, LCLs, pancreatic islests, etc).

Notes from meeting:

* Programs uses:
  + tensorqtl (python package) for conditional analysis (because it gets standard errors values that R doesn’t)
  + Colocalization part is in R
  + So, part python and part R
* Interesting questions:
  + Is fine mapping suitable? (suzie)
* What do to:
  + Do Conditional analysis
  + Pick a gene and look at colocalization
  + Analyse this data
  + Step 0: wait until lecture sends us the location of the data we need to analyse for tensorqtl
  + Step 1: Perform normal eQTL analysis for SIGLEC14 gene, testing all common (MAF > 1%) variants in +/- 1Mb window around the promoter of the gene.
    - (Like what we did with the CD14 when we got Manhattan plot)
  + Step 2: Identify the most strongly associated variant (lead variant)
  + Step 3. Perform conditional eQTL analysis for SIGLEC14 by adding the lead variant as a covariate into the model.
  + Step 4: Repeat this until no significant (p< 1e-5) associations remain.
  + Step 5: Identify all-but-one conditionally indepedent summary statistics. If in Step 4 you identify 3 conditionally independent signals, then at this step you need to perform conditioning three times:
    - for Signal 1, condition on Signals 2 and 3 (add both lead SNPs as covariates).
    - for Signal 2, condition on Signals 1 and 3
    - for Signal 3, condition on Signals 1 and 2.
* Orga:
  + If questions write him on slack and also post progress there
    - Meetings will also be there
  + Needs to be finished until: something like 12th of June
    - Presentation of results in class
    - And short report what we did
  + Goal: does the project analysis work or not?
    - Can we do ??? when one is fine mapped and other is conditional analyzed
    - Reason for this is because conditional analyzation is very expensive
    - Lecture cares about result. Answer would be useful for very projects. But could be that it cannot be done.

Effect of two variants:

* X-axes = effect on gene expression

