Diary – Variant to Gene Mapping analysis (since 20/10/2023)

**20/10/2023**

Updates scripts for GTEx colocalisation. Worked on the 004\_concat\_results.R

Created a .xlsx file in the src/report/Variant\_to\_Gene\_Tables.xlsx; for tables on V2G analysis. Put a table with colocalisation data: locus, tissue, N\_gene\_sign/N\_gene\_tot

I still have to modify files for eQTLGen colocalisation; still to obtain GTExV8 files for Colon\_Transverse and Colon\_Sigmoid

Did colocalisation for ‘Stomach’ and ‘Small\_Intestine\_Terminal\_Ileum’ **\***

**23/10/2023**

We have OK for using U-BIOPRED eQTL data with genotyped data.

We have OK for using UBC Lung eQTL data.

There was an error in the script to run colocalisation with GTExV8. So I had to run colocalisation again for ‘Stomach’ and ‘Small\_Intestine\_Terminal\_Ileum’ **\***

After discussion with team, I do colocalisation only if the eQTL data for the tissue-gene-credset region contains significant association, aka pvalue <= 5x10-6. 🡪 Updated 003\_run\_coloc\_susie\_GTEx.R to integrate this step.

Update scripts with some quality checks as well on the number of genes analysed, analysed by colo, analysed by coloc.susie

**24/10/2023**

Add additional checks in the colocalisation with GTExV8.

Removed the additional checks line form the Var\_to\_Gene\_pipeline.sh and some from 003\_run\_coloc\_susie\_GTExV8.R

Submit coloc for GTExV8 ‘Lung’, ‘Small Intestine Terminal Ileum’, ‘Stomach’, ‘Esophagus Muscularis’.

**30/10/2023**

Re-read the report as it is up today, and updated it a little.

Put a new check to find if all the genes for each tissue have been analysed.

Run colocalisation for 'Esophagus\_Gastroesophageal\_Junction', ‘Artery\_Tibial’, ‘Artery\_Coronary’

STILL NEED TO RUN COLOCALISATION FOR ‘ARTERY\_AORTA’

STILL NEED TO CREATE eQTL FILES FOR COLON\_TRANSVERSE AND COLON\_SIGMOID.

NEED TO CODE 004\_concat\_coloc\_results.R FOR COLOC.SUSIE RESULTS

**31/10/23**

Run colocalisation for 'Artery\_Aorta’

Started working on the liftOver of eQTL data, for Colon Transverse and Colon Sigmoid, with script 000\_liftover\_b38\_to\_b37\_GTExV8.sh, based on Chiara’s script. I wanted to use liftOverPlink, but I can’t with bed file only. Excursus: needed to modify the exe file of liftOverPlink for python3 – print command wants parenthesis for the argument. I needed to download liftOver, apparently not installed in ALICE3.

So, Chiara explained me that the script 000\_liftover\_b38\_to\_b37\_GTExV8.sh does the liftOver on all individuals of GTExV8, meanwhile I am interested in European ancestry individuals. So, I have to start from a different set of GTExV8 .parquet data, as downloaded from the website (<https://www.gtexportal.org/home/downloads/adult-gtex#qtl>) and present in ALICE folder: /data/gen1/ACEI/colocalisation\_datasets/eQTL/GTeX

I have to 1)convert hg38 .parquet file into hg38 .gz file; 2)liftOver hg38 .gz file into hg19 .gz file. In this way, I will obtain the same data Kayesha did for the other tissues (/data/gen1/ACEI/colocalisation\_datasets/eQTL/GTeX/${tissue}.v8.EUR.allpairs.chr${chr}.hg19.txt.gz).

So, I am now looking at Kayesha’s scripts.

Created 000A\_submit\_eqtl\_gtex\_extraction.sh: ok, problem with chromosome X (segmentation issue ?, need to understand)

Other scripts are: 000A\_eqtl\_gtex\_extraction.R; 000B\_eqtl\_gtex\_liftover.sh; 000C\_eqtl\_gtex\_conversion.R

STILL NEED TO ADD THE SCRIPTS ON THE REPORT AND TO RUN THEM.