System genetics final exercise

## Submitted by: Guy Shur 311133110, Dror Hadas 302678263

# Background & Data

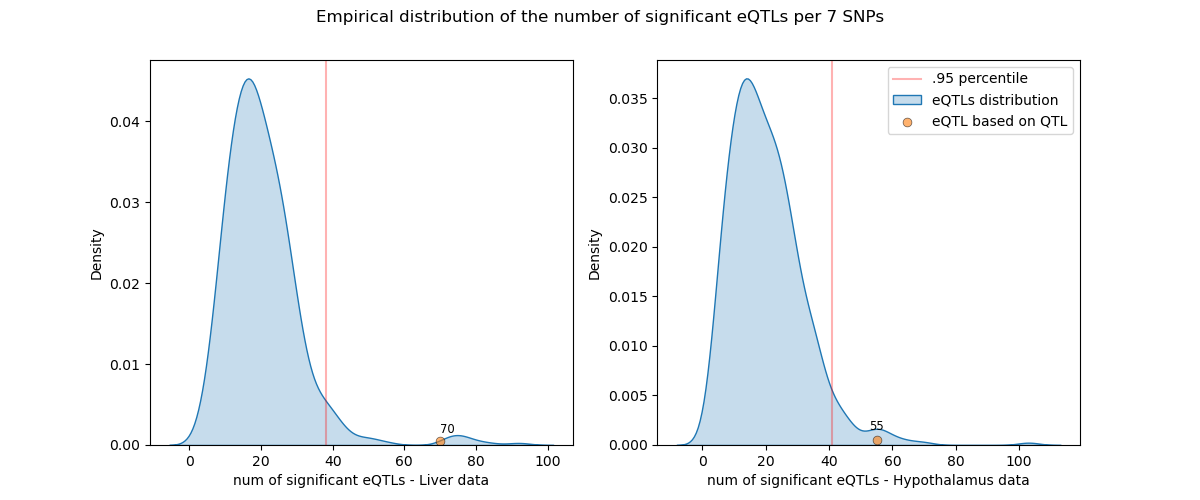
# Q.5 – Combining eQTL and QTL results

Since QTLs represent a possible relation between a SNP and a trait, a phenotype, we hypothesized that an intermediate later of connections could be hidden in the eQTL data. A QTL may effect a phenotype by effecting a related expression patten, that is via an eQTL.

To test this hypothesis, we have used the significant SNPs from the QTL analysis (see previous section) in order to narrow down the eQTL dataset. The QTL analysis yielded 9 SNP-phenotype pairs that originated from 7 SNPs and 3 phenotypes. Upon narrowing the eQTL (post FDR) to only the significant SNPs from the QTL analysis, 70 & 55 eQTL were found significant in the Liver and Hypothalamus datasets accordingly.

To determine whether the number of significant eQTLs is higher than expected, we have generated 500 random datasets of eQTL data where each dataset is comprised of only 7 random SNPs (In contrast to all possible SNPs) and all expression data. We then checked for each such dataset how many eQTLs were significant and created an empirical distribution of the number of significant eQTLs per a dataset of 7 SNPs.

Our next step was plotting a kernel density estimation of these distribution and check whether 70 and 55 are higher than the 95th Percentile.



In both datasets limiting the eQTL analysis to only QTL significant SNPs yields more significant eQTLs than expected by random. In both datasets the mean number of significant eQTLs is ~20, while 70, 55 are much higher than then 95th percentile.

The described analysis above, was performed on eQTL data that was correct for FDR when considering the number of statistical tests derived by all SNPs and not only those who were proven significant in the QTL analysis. It is reasonable to repeat the analysis, while preforming the FDR correction separately for each random dataset and see if it increases the statistical power of the eQTL analysis.

Chart

Description automatically generated

As expected, narrowing the eQTL analysis to only QTL data and then preforming FDR yields more significant eQTLs then first preforming FDR while considering a larger number of tests. However it can be seen that while the trend is strictly maintained in the liver data, in the hypothalamus data the number of significant eQTLs (77) is closer to the 95th percentile of the empirical distribution.

Overall, the strategy of narrowing eQTL analysis to only QTL significant SNPs seems reasonable and productive and could be used when computational resources are a bottleneck. As mentioned above, from our perspective it seems to have a biological incentive as well since a SNP might effect a phenotype via effecting the expression of related genes.