



Project: Metabolic QTLs and Metabolic Networks

(Total 80 Points)

Estimated Time: 2.5 days in Groups of 2

Objective: This project aims to perform genomic analysis, including kinship calculation, SNP-metabolite association discovery (mQTL), visualization, metabolic network construction, network visualization, genetic variant annotation, regional plots, and inflation factor calculation.

You will use the following datasets in this project:

1. [Dataset of 156 Qataris](#)
2. [Metabolites](#)

Task 1: Compute Kinship using SNPRelate and GENESIS (10 points):

- a. Compute kinship using the SNPRelate and GENESIS package in R. You can use the IBD family of functions but note that you need to transform the plink format to GDS format using the function `snpgdsPED2GDS` that takes the ped and map files.
- b. Report the number of individuals who have a kinship > 0.1

Task 2: Compute mQTLs with Mixed Models (15 points):

- a. Compute mQTLs using all SNPs and all metabolites to identify associations and report their p-values. You will need to use mixed models to include kinship.
- b. GENESIS provides a `fitNullModel` function that computes residuals of metabolites from the covariates including kinship and a `assocTestSingle` function that computes the mQTLs.
- c. Include the first three principal components in the covariates.
- d. Report significant SNP-Metabolite associations until $p < 0.01$ in an Excel sheet, showing beta values, standard error, effect allele, degree of freedom, and p values. (all values are the direct output of `assocTestSingle`).
- e. Report heritability from the `varCompCI` function in GENESIS.

Task 3: Inflation factor calculation (10 points): (Reference slide given)

Calculate and report the inflation factor for each metabolite in the mQTL analysis and report them and their average.

Task 4: Manhattan Plot (10 points):

Plot a Manhattan plot for ALL associations (significant or not). Hint: Make sure the chromosomes have different colors and a significance line is included. You can use (qqman) or any other library in R.

Task 5: Metabolic Networks (15 points):

- a. Correct metabolites for covariates and kinship using the residuals obtained from polygenic function.
- b. Use the corrected metabolites as input to the GeneNet package to get significant pairs of partially correlated metabolites and write all pairs to a CSV file.
- c. Use Cytoscape to draw the metabolic networks. Add names to nodes.

Task 6: Annotate Significant SNPs (10 points):

Annotate 10 SNPs that have the top significant associations with metabolite using seattle seq annotations or Annovar tool. Append the annotations to the significant associations reported in task 2(d).

Task 7: Regional plots using SNIPA (10 points):

Plot the regional plots of the 5 most significant associations using SNIPA or the regional plot of the nearest SNP found in SNIPA. This will use the metabolic values for metabolites associated with those SNPs in SNIPA

References:

1. [PLINK official documentation](#)
2. [Dataset of 156 Qataris](#)
3. [Metabolites](#)