

# eQTL Exercise

## Understanding the data

**Task 1: Take a look at the sub\_geno.tab, sub\_expr.tab and design.tab files.**

- What do the -1,0,1,2 values represent in the sub\_geno.tab file?
- What is stored in the sub\_expr.tab file and what has been done with this data?
- What information is stored in the design.txt file?

## Task 2: Genotype data

- Calculate the number of missing genotypes for each SNP across all individuals.
- Calculate the minor allele frequency (MAF) for all SNPs across all individuals.
- Filter our SNPs that have missing genotypes or a  $MAF < 0.05$  and use the filtered snps for the rest of the exercise.
- Calculate the MAF for africans and non-africans separately. Is there a difference?

## Task 3: Gene expression profiles

- Plot the distribution of expression levels across all samples for the ENSG00000172404.4 gene
- Plot the expression levels of ENSG00000172404.4 against the genotypes of snp\_22\_41256802 and snp\_22\_45782142

## Linear regression of genotype and phenotype with and without covariates

**Task 4: Do a linear regression of all sample genotypes on sample gene expression:**

- For snp\_22\_41256802 on ENSG00000172404.4
- For snp\_22\_45782142 on ENSG00000172404.4
- Make sense of the results (Understand what the values represent)

**Task 5: Do a linear regression for snp\_22\_43336231 on ENSG00000100266.11**

- Without covariates
- Using the genotype PCs from pc\_cvrt.tab as covariates
- Separately for african and non-africans without covariates
- Make sense of the results. Some plots may help

## Multiple testing

**Task 6: Do a linear regression on 1st snp on 1st gene, 2nd snp on 2nd gene etc.**

- Create a matrix containing the gene\_id, snp\_id, effect size, t.value and p.value.
- Do a multiple testing correction on the resulting p.values.
- Do the same but now include the genotype PCs from pc\_cvrt.tab as covariates. extra: Plot significant eQTLs (genotype vs expression)

## Matrix eQTL

**Task 7: Use the Matrix\_eQTL\_main function to do eQTL analysis on the data.**

- You will need the sample\_geno.pos and sample\_expr.pos for this. What do these files contain?
- Read the Matrix\_eQTL\_main manual and decide on what settings to use.
- Run your Matrix eQTL analysis and compare the results to the results from the linear models.