# Assignment 1: Advanced Computational Biology

All code used to answer the questions are at the end this document.

## Question 1

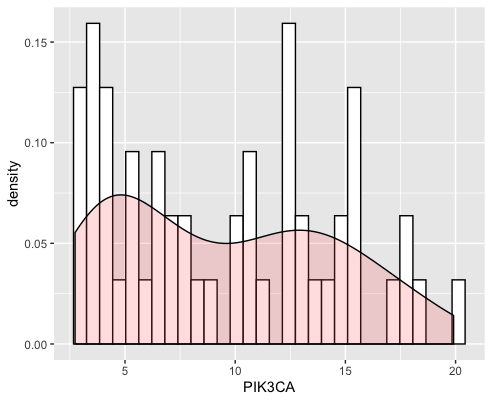
a) SNPs below the 10 -8 cut off include rs4977264 (1.063495e-11) and rs113505981 (4.725661e-14). It was found that only rs4977264 complied with Hardy-Weinberg equilibrium (Chi-squared test p-value = 0.6181).

Plotted are the distribution of gene expression levels across samples for gene PIK3CA, hence illustrating identified associations.

A screenshot of a cell phone

Description automatically generated

Base R:

Ggplot2:

A screenshot of a cell phone

Description automatically generatedBoxplot:

b)

A screenshot of text

Description automatically generated

We can see (above) that there is a significant association between the SNP (rs4977264) and PIK3CA expression (p-value: < 2.2e-16), but also between colon, kidney, lung, oesophagus, and pancreas expression. The lower p-value when using different covariates may indicate that the regression describes the data more accurately.

A linear regression and a fixed effect model were built taking into account the tissue of origin. The two models are significantly different (p-value: < 0.002935; 0.05 threshold), so there is a significant association between the selected SNP and PIK3CA expression after taking into account the site where measurements were performed. This result is similar in comparison to the results in the previous findings. Mixed effect models take into account various random effects that may not be causal to the trait in question. The lower p-value when using different covariates may indicate that the regression describes the data more accurately.

## Question 2

a) There are 124 significant SNPs in total which are below the threshold; 121 for chromosome 9 and 3 for chromosome 17.

b) Manhattan plot chromosome 9

A screenshot of a cell phone

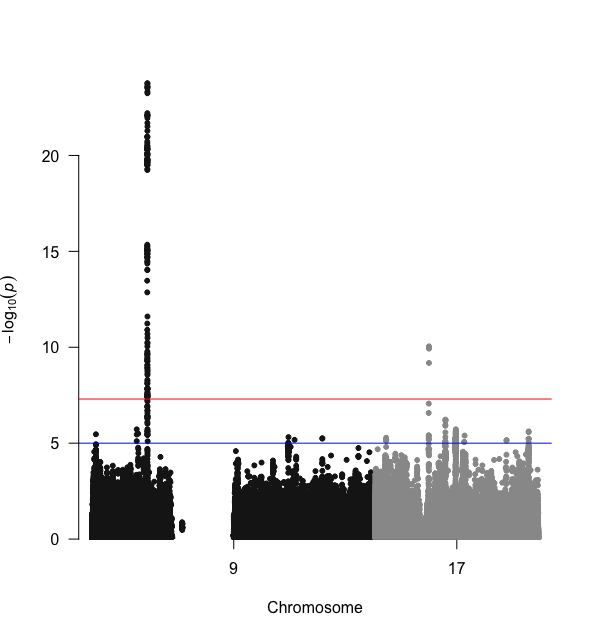
Description automatically generated

Manhattan plot chromosome 17

A screenshot of a cell phone

Description automatically generated

Manhattan plot for all the data using the qqman package



The plot supports the above results showing a peak at chr 17 (3points) and a

larger, more dense peak at chr 9 (121 points). The X axis shows that the position on the

chromosome where these significant SNPs are located are at similar position. The Y

axis tells how much it is associated with a trait, these vary in range surpassing the threshold.

Noticeable also is that there is a gap in the positions of chromosome 9 where there are no SNP records at all.

## Question 3

a)

The topmost significant SNP from chromosome 9 is SNP rs3849943.

The topmost significant SNP from chromosome 17 is SNP rs35714695.

For Chr 9 one gene is affected, C9orf72 and the tissues affected are as seen in the table below. For Chr 17 two genes, TMEM97 and POLDIP2 are affected and the tissues affected are as seen in the table below.

A screenshot of a cell phone

Description automatically generated

## A screenshot of a computer Description automatically generated

b)

For Chr 9 one gene is affected, C9orf72 and the tissues affected are as seen

in the table above. For Chr 17 two genes, TMEM97 and POLDIP2 are affected and

the tissues affected are as seen in the table above.

Chr 9: All SNP.ID rs3849943, gene C9orf72 have Cis-eQTLs (number = 18)

Chr 17: All SNP.ID rs35714695, gene TMEM97 have Cis-eQTLs (number = 1)

All SNP.ID rs35714695, gene POLDIP2 have Cis-eQTLs (number = 2)

A cis eQTL is an SNP showing association with gene expression located close by (normally around 1Mb) of the gene that it influences. In contrast a trans eQTL is a SNP located elsewhere in the genome, a distance from the gene whose expression it is altering. GTEx IGV Browser was used to explore the location of the SNP in comparison to the transcription start sites of the genes, hence evaluate the relative distance. Also, an online eQTL browser was also used as another source of information. (Lonsdale *et al.*, 2013; Westra *et al.*, 2013). In addition, gene expression signatures are cell-type specific, and therefore regulatory control of expression may also be cell-type dependent. For example, significant tissue specificity has been reported for multiple cis eQTLs.

c)

POLDIP2

The brain cis-eQTL effect for SNPs in this locus on POLDIP2 suggests that

POLDIP2 could be the causal gene in this locus. Another overlap was observed

in the SARM1 locus where rs35714695 and its proxies had the strongest

exon-level cis-eQTL effect on POLDIP2 in multiple brain tissues (Van Rheenen *et al.*, 2016).

TMEM97

TMEM97 ligands bind to S2R and has a pharmacologic profile the same as that of S2R. Over the past few decades, sigma receptors (SRs), including sigma 1 and sigma 2 receptor subtypes (S1R and S2R, respectively) have been widely associated with aging- and mitochondria-associated disorders, such as Parkinson’s and Alzheimer’s disease, multiple sclerosis and amyotrophic lateral sclerosis. However, the specific role played by this orphan receptor family in cell biology has yet to be clarified. The 3D structure of TMEM97/S2R would provide understanding of the biological functions and mechanisms (Mavlyutov *et al.*, 2015; Tesei *et al.*, 2018).

C9orf72

The main functional disease mechanisms proposed have been toxic gain of function from C9orf72 repeat RNA or from dipeptide repeat proteins produced by repeat associated non-ATG translation and the loss of function of the C9orf72 protein. More than a few hundred repeats represent a risk for ALS and FTD.The discovery that repeat expansions in the C9orf72 gene are a frequent cause of amyotrophic lateral sclerosis (ALS) has been important in understanding the pathogenic mechanisms of the disease. Several downstream processes across a range of cellular functions have also been implicated (Balendra and Isaacs, 2018; Leko *et al.*, 2019).

## Code

## A screenshot of a computer screen Description automatically generatedQuestion 1 a)

A screenshot of a computer screen

Description automatically generated

## Question 1 b)

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer screen

Description automatically generated

## Question 2

A screenshot of a computer screen

Description automatically generated

## Question 3

A screenshot of a computer screen

Description automatically generated

## References

Balendra, R. and Isaacs, A. M. (2018) ‘C9orf72-mediated ALS and FTD: multiple pathways to disease’, *Nature Reviews Neurology*. Nature Publishing Group, pp. 544–558. doi: 10.1038/s41582-018-0047-2.

Leko, M. B. *et al.* (2019) ‘Molecular mechanisms of neurodegeneration related to C9orf72 hexanucleotide repeat expansion’, *Behavioural Neurology*. Hindawi Limited. doi: 10.1155/2019/2909168.

Lonsdale, J. *et al.* (2013) ‘The Genotype-Tissue Expression (GTEx) project’, *Nature Genetics*, pp. 580–585. doi: 10.1038/ng.2653.

Mavlyutov, T. A. *et al.* (2015) ‘Role of the Sigma-1 receptor in Amyotrophic Lateral Sclerosis (ALS)’, *Journal of Pharmacological Sciences*. Japanese Pharmacological Society, pp. 10–16. doi: 10.1016/j.jphs.2014.12.013.

Van Rheenen, W. *et al.* (2016) ‘Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis’, *Nature Genetics*. Nature Publishing Group, 48(9), pp. 1043–1048. doi: 10.1038/ng.3622.

Tesei, A. *et al.* (2018) ‘Sigma receptors as endoplasmic reticulum stress “gatekeepers” and their modulators as emerging new weapons in the fight against cancer’, *Frontiers in Pharmacology*. Frontiers Media S.A. doi: 10.3389/fphar.2018.00711.

Westra, H. J. *et al.* (2013) ‘Systematic identification of trans eQTLs as putative drivers of known disease associations’, *Nature Genetics*, 45(10), pp. 1238–1243. doi: 10.1038/ng.2756.