**UCL DIVISION OF BIOSCIENCES**

***IN COURSE ESSAY SUBMISSION AND FEEDBACK FORM (LEVEL 7)***

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CANDIDATE NO.** | **D** | **V** | **Y** | **M** | **9** |  |  |  | **Submission Date:** | **04/02/2020** |
| **Module Code:** | **BIOL0050** | | | | | | | | **Word Count:** | **1100** |
| **Title/Type of Coursework:** | **Assignment 1: Advanced Computational Biology** | | | | | | | | **Turnitin Submission Receipt** |  |

|  |  |  |
| --- | --- | --- |
| **Have you been diagnosed as having Dyslexia by the UCL Dyslexia Centre?** | **Y** | **N** |

**SELF-ASSESSMENT (FOR STUDENTS TO COMPLETE)**

|  |
| --- |
| **Based on previous feedback, try to identify specific aspects of this work that you would like additional feedback on.** |

**FEEDBACK (FOR MARKERS TO COMPLETE)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Marker Name:** |  | **Date:** |  |
| 1. **What the student did well.** | | | |
|  | | | |
| 1. **What the student did not do well.** | | | |
|  | | | |
| 1. **Action: How might the student improve?** | | | |
|  | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **First Mark (%)** |  | **Second Mark (%)** |  | **Final Mark (%)** |  |

**NOTE:**

All assessed coursework is subjected to review by a Second Marker or the Course Organiser. The mark given here remains provisional until confirmed at the Examination Board.

# Assignment 1: Advanced Computational Biology

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

## Question 1

a)

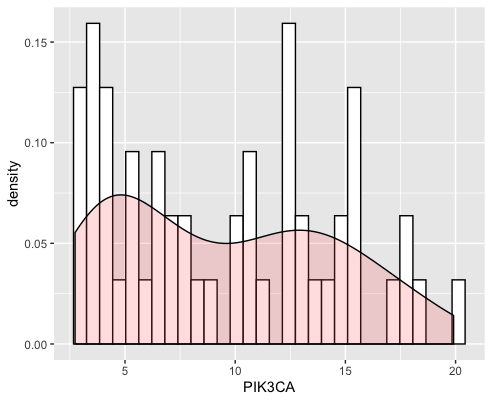
Quality control procedures can be utilised to correct for population effects, genotyping errors and potential deviations from common assumptions. Major quality control steps were used to correct for errors in this dataset, as observed in the code.

SNPs below the 10 -8 cut off include rs4977264 (1.063e-11) and rs113505981 (4.726e-14). It was found that only rs4977264 complied with Hardy-Weinberg equilibrium (Chi-squared test p-value = 0.6181).

Plotted below are the distribution of gene expression levels across samples for the gene PIK3CA, hence illustrating identified associations. A boxplot is also used to investigate the SNP rs4977264 variant effect on the gene PIK3CA and the expression levels by genotype.

A screenshot of a cell phone

Description automatically generated

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 where the expression measurement of PIK3CA was performed for each individual: e.g. 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 mixed 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. Mixed effect models take into account various random effects that may not be causal to the trait in question. This result is similar in comparison to the results in the previous findings.

## Question 2

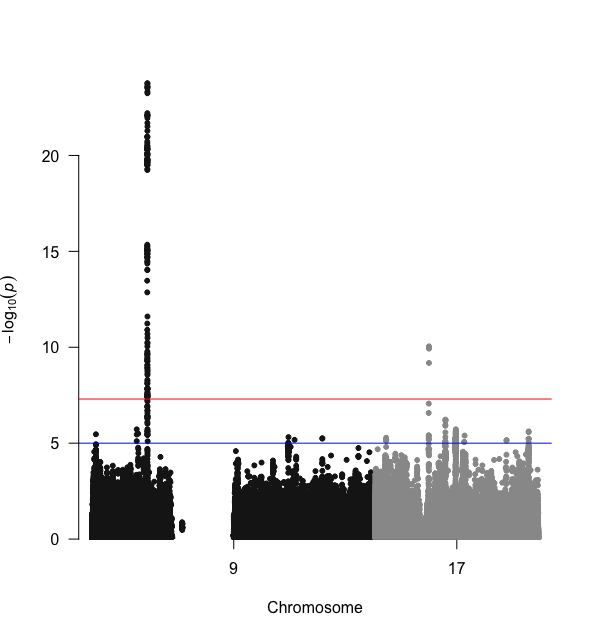
1. There are 124 significant SNPs in total which are below the threshold; 121 for chromosome 9 and 3 for chromosome 17.

A screenshot of a cell phone

Description automatically generatedb) Manhattan plot chromosome 9

A screenshot of a cell phone

Description automatically generatedManhattan plot chromosome 17

Manhattan plot for all the data using the qqman package

Each point represents -log10(P) of association with trait, for an SNP, ordered by chromosome. The threshold of significance is indicated as a horizontal red line. The plot supports the above results showing a significant peak at chromosome 17 at locus 17q11.2 (3 SNPs) and a larger, denser peak at chromosome 9 at locus 9p21.2 (121 SNPs). The Y axis tells how much it is associated with a trait; these vary in range and surpasses the threshold at two chromosome locations. Noticeable also is that there is a large gap in the positions of chromosome 9 where there are no SNP records at all. This is may be due to unsequenced chromosome regions; they may also be duplicons, part of chromosome 9 low-copy repeats pericentromeric region (LCR9-1 to LCR9-4) which are part of human genome duplicated segments which were previously uncharacterised (Paulis et al., 2006). Many associated variants are unlikely to be causal due to neighbouring markers which may be inherited together causing linkage disequilibrium between two markers. There may also be heterogeneity in ALS risk loci across different ethnicities. A study of ALS in a Han Chinese population identiﬁed two loci of genome-wide signiﬁcance, however these loci were not replicated in populations of European ancestry (Deng et al., 2013). The Manhattan plot was produced using the qqman package in R.

At locus 9p21.2, 18 SNPs close to Corf72 gene lie

above the red line (most signiﬁcant SNP is rs3849943 with P¼7.69 ×10

29

). Locus 17q11.2 shows SNP rs34517613 (P¼1.11 ×10

28

) to be signiﬁcantly asso-

ciated. SNP rs1788776 at 18q11.2 is very close to the threshold of signiﬁcance with P¼7.67 ×10

28

.

## Question 3

a)

The topmost significant SNP from chromosome 9, in the study 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 directly below.

A screenshot of a computer

Description automatically generatedA screenshot of a cell phone

Description automatically generated

b)

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 on the locus in 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 tissues including the brain, as illustrated in the table above (Van Rheenen et al., 2016).

TMEM97

TMEM97 ligands bind to S2R and has a similar pharmacologic profile 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, Alzheimer’s disease 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 greater 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, the loss of function of the C9orf72 protein and from dipeptide repeat proteins produced by repeat associated non-ATG translation. Greater than a few hundred repeats represent a risk for FTD and ALS. The discovery that repeat expansions in the C9orf72 gene are a frequent cause of amyotrophic lateral sclerosis has been important in understanding the pathogenic mechanisms of the disease. The advantage of the identification and specification of translating risk-SNPs to notable genes is that actions towards drug discovery may be accelerated. Several downstream processes across a range of cellular functions have also been implicated (Balendra and Isaacs, 2018; Leko *et al.*, 2019).

## Code

## Question 1 a)

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a cell phone

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.

Deng, M. et al. (2013) ‘Genome-wide association analyses in Han Chinese identify two new susceptibility loci for amyotrophic lateral sclerosis’, Nature Genetics. Nature Publishing Group, 45(6), pp. 697–700. doi: 10.1038/ng.2627.

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.

Paulis, M. et al. (2006) ‘A set of duplicons on human chromosome 9 is involved in the origin of a supernumerary marker chromosome’, Genomics, 87(6), pp. 747–757. doi: 10.1016/j.ygeno.2006.02.014.

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.

**Faculty of Life Sciences Level 7 (Postgraduate or Integrated Undergraduate Masters) Essay Marking Guidance**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Marks | 0-39 | 40-49 | 50-59 | 60-69 | 70-85 | 86-100 |
| Fail/Condonable Fail | | Condonable Fail | Pass/2ii | Merit/2i | Distinction/First | Distinction/First |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Knowledge and understanding of field**  *e.g. analysis and synthesis; deploying logical argument supported by evidence; focus on topic; drawing conclusions; ability to communicate ideas or material to diverse audiences* | Demonstrates little knowledge of the field.  Demonstrates significant weaknesses in the knowledge base, and/or simply reproduces knowledge without evidence of understanding.  Shows little or no critical ability.  Poor, inconsistent analysis. Use 0-15 where no answer or makes a single relevant point or a few disconnected sentences/ images. | Demonstrates knowledge of the field and awareness of current evidence and its interpretation, but with some notable weaknesses.  Lacks knowledge and understanding of some key areas.  Offers some appropriate analysis, but with some significant inconsistencies which affect the soundness of argument and/or conclusions.  Demonstrates very limited critical ability. | Demonstrates a sound knowledge and understanding of material within the field and is up-to-date with current developments.  Demonstrates relevant and sound analysis of material presented with some critical evaluation.  Is able to analyse complex questions and make appropriate judgements. May lack clear focus on the question in parts.  Able to communicate arguments, evidence and conclusions to specialist and non-specialist audiences. | Produces work with a clear focus on the question throughout.  Demonstrates a systematic knowledge, understanding and critical awareness of current problems and/or new insights, much of which are at the forefront of the field and informed by it.  Is able to evaluate methodologies critically.  Is able to deal with complex problems both systematically and creatively and make sound judgements in the absence of complete data.  Consistently able to communicate arguments, evidence and conclusions to specialist and non-specialist audiences. | Produces work reflecting excellent understanding of the question.  Displays mastery of a complex and specialised area with notable critical awareness of current problems and/or new insights at forefront of field.  Shows excellent ability to evaluate methodologies critically.  Deals with complex problems systematically and creatively, making excellent judgements.  Consistently able to communicate high-level arguments, evidence and conclusions to diverse audiences. | This work meets and often exceeds the standard for distinction/first class, as described in the 70-85 band, across *all* sub-categories of criteria.  The ideas and analyses presented are of publishable quality (with only very minor amendments) and would be likely to receive that judgement if submitted to a peer-reviewed journal.  Work is of such a quality that the student is clearly highly capable of doctoral research in the discipline and would be prioritized for a PhD studentship or DTP placement. |
| **Research and future enquiry**  *e.g. framing and creating further question or new hypotheses;*  *suggesting appropriate methods for gathering further evidence; awareness of methodological benefits/ limitations;*  *critical analysis of evidence* | Lacks any understanding of how established techniques of research and enquiry are used to create and interpret knowledge in the field.  Little or no skill demonstrated in understanding or selecting techniques applicable to the specific question. | Lacks sufficient understanding of how established techniques of research and enquiry are used to create and interpret knowledge in the field  Demonstrates some skill in selected techniques and/or approaches applicable to the specific question but with significant areas of weakness. | Shows some originality in the application of knowledge, and some understanding of how established techniques of research and enquiry are used to create and interpret knowledge in the discipline  Demonstrates understanding of, and skills in, selected techniques/ approaches applicable to the specific question | Shows originality in the application of knowledge, and a good understanding of how knowledge is created and interpreted in the discipline.  Displays a comprehensive understanding of, and skills in, techniques/approaches applicable to the specific question.  Able to identify gaps in current knowledge and formulate hypotheses or research proposals | Shows originality in application of knowledge, and excellent grasp of how knowledge is created and interpreted in the discipline.  Displays exceptional grasp of a range of techniques applicable to the specific question.  Able to identify gaps in current knowledge and formulate original hypotheses or research proposals |